BAYESIAN ANALYSIS OF ARMA PROCESSES: COMPLETE SAMPLING BASED INFERENCE UNDER FULL LIKELIHOODS

by

John Marriott

N. Ravishanker

Alan E. Gelfand

TECHNICAL REPORT No. 471

June 24, 1993

Prepared Under Contract

N00014-92-J-1264 (NR-042-267)

FOR THE OFFICE OF NAVAL RESEARCH

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA 94305-4065

BAYESIAN ANALYSIS OF ARMA PROCESSES: COMPLETE SAMPLING BASED INFERENCE UNDER FULL LIKELIHOODS

by

John Marriott

N. Ravishanker

Alan E. Gelfand

TECHNICAL REPORT No. 471

June 24, 1993

Prepared Under Contract

N00014-92-J-1264 (NR-042-267)

FOR THE OFFICE OF NAVAL RESEARCH

Professor Herbert Solomon, Project Director

Approved for public release; distribution unlimited

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA 94305-4065

BAYESIAN ANALYSIS OF ARMA PROCESSES: COMPLETE SAMPLING BASED INFERENCE UNDER FULL LIKELIHOODS

John Marriott, Nalini Ravishanker, Alan Gelfand

and

Jeffrey Pai*

Abstract

For a general stationary and invertible ARMA (p,q) process, we show how to carry out a fully Bayesian analysis. Our approach is through the use of sampling based methods involving three novel aspects. First the constraints on the parameter space arising from the stationarity and invertibility conditions are handled by a convenient reparametrization to all of Euclidean (p+q)-space. Second, required sampling is facilitated by the introduction of latent variables which, though increasing the dimensionality of the problem, greatly simplifies the evaluation of the likelihood. Third, the particular sampling based approach used is a Markov chain Monte Carlo method which is a hybrid of the Gibbs sampler and the Metropolis algorithm. We also briefly show how straightforwardly the sampling based approach accommodates missing observations, outlier detection, prediction and model determination. Finally we illustrate the approach with two examples.

Keywords: ***Gibbs sampler***, ***invertibility***, ***latent variables***, ***Metropolis algorithm***, ***missing values***, ***outliers***, ***prediction***, ***stationarity***.

1. Introduction

Autoregressive moving average models constitute a broad class of parsimonious time series processes which are useful in describing a wide variety of time series. In particular, the process $\{z_t\}$ is said to be an autoregressive moving average process of order p, q, ARMA (p, q), with mean μ, if it is generated by

$$\phi(B)(z_t-\mu) = \theta(B)\epsilon_t \tag{1}$$

where B is the backward shift operator, $B^k z_t = z_{t-k}$, $\{\epsilon_t\}$ is a Gaussian white noise process with variance σ^2, $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ and $\theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$ are polynomials in B of degrees p and q respectively.

For the process to be stationary the roots of $\phi(B) = 0$ must lie outside the unit circle and if the roots of $\theta(B) = 0$ also lie outside the unit circle the process is said to be invertible, in which case there is a unique model corresponding to the likelihood function. It may be argued (Box and Jenkins, 1976) that, in forecasting, a nonstationary process is undesirable and a noninvertible process is meaningless.

If the stationarity and invertibility conditions are to be satisfied, the parameter vectors, $\boldsymbol{\phi} = (\phi_1,\cdots,\phi_p)$ and $\boldsymbol{\theta} = (\theta_1,\cdots,\theta_q)$, are constrained to lie in regions C_p and C_q, respectively, corresponding to the polynomial operator root conditions. Allowable values of $(\boldsymbol{\phi}, \boldsymbol{\theta})$ then lie in $C_p \times C_q$, the forms of which are simple to identify for $p \leq 2$, $q \leq 2$. However for $k > 2$ the form of C_k becomes complicated and for $k > 4$ the polynomial equations $\phi(B) = 0$ and $\theta(B) = 0$ cannot, in general, be solved analytically.

In a Bayesian setting, for a stationary invertible ARMA (p,q) time series model of the form (1), the region $C_p \times C_q$ determines the ranges of integration for obtaining joint and marginal distributions of the parameters and for evaluating posterior expected values. Historically, Bayesian analysis of these models ignores this region in order to obtain convenient distributional results for the posterior densities. See, e.g., Zellner (1971), Box and Jenkins (1976), Broemeling (1985), Broemeling and Shaaraway (1988). Only Monahan

(1983) and Marriott and Smith (1992) have restricted the analysis to $C_p \times C_q$ employing analytic numerical integration techniques. Monahan's calculations used fixed point quadrature rules for $p+q \leq 2$ while Marriott and Smith extended these results using the Gauss–Hermite rules described in Smith et. al. (1987). Generally, to accurately perform such numerical integration requires specialist software and expertise and typically, for a model with unknown μ and σ^2, becomes infeasible for $p+q \geq 4$.

We propose the use of sampling based methods to obtain desired marginal posteriors. In particular, the introduction of the Gibbs sampler as a tool for Bayesian calculations in Gelfand and Smith (1990) has spurred considerable interest in Markov chain Monte Carlo methods. We use a variant of the Gibbs sampler described in Müller (1991). Recently Chib (1991) and Chib and Greenberg (1992) have utilized the Gibbs sampler for Bayesian analysis of AR and MA processes. Their work differs from ours in several ways. Most notably, in the first of these papers the conditional likelihood is used while in the second a computationally expensive form of the unconditional likelihood is used. In neither paper is the full ARMA model handled. Proposed sampling from complete conditionals for ϕ and θ requires a "double rejection" method. In our early work we abandoned such sampling as terribly inefficient for larger p and q. We also note that McCulloch and Tsay (1991) have employed the Gibbs sampler for a Bayesian analysis of autoregressive processes again ignoring stationarity restrictions. Carlin, Polson and Stoffer (1992) use the Gibbs sampler for Bayesian analysis of first order dynamic models also without concern for stationarity.

The outline of the paper then is as follows. In Section 2 we give the ARMA likelihoods in a form well suited for our Monte Carlo method. In Section 3 we obtain the complete conditional distributions required for the Gibbs sampler. Section 4 provides a useful transformation to enable us to handle the stationarity and invertibility constraints. Section 5 describes a modified Gibbs sampler which works well for time series models with

Gaussian error. Section 6, 7 and 8 briefly deal with the issues of missing data, prediction and model determination. Finally Section 9 analyzes two interesting data sets.

2. Likelihood forms

In Section 2.1 we note that the likelihood for the general ARMA (p,q) model can be written in a relatively simple form by conditioning upon the unobserved history of the process. This form is inexpensive to evaluate making it attractive for use with the Gibbs sampler. In Section 2.2 we observe that for the pure autoregressive process, AR(p), we can marginalize over this history resulting in a lower dimensional likelihood which can still be evaluated cheaply.

2.1 The general ARMA (p,q) likelihood

For the general ARMA (p,q) model in (1) the form of the exact likelihood is known (Newbold, 1974). While in theory it is possible to work with this form, its evaluation, unless the sample size is very small, is computationally very expensive rendering it infeasible for use with a Monte Carlo approach. A form for the likelihood which is well suited for sampling based techniques can be developed by introducing latent variables (parameters). Dimensionality of the parameter space is increased in exchange for a manageable likelihood.

Letting $y_t = z_t - \mu$, (1) can be written as

$$y_t = \sum_{j=1}^{p} \phi_j y_{t-j} - \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t. \qquad (2)$$

Hence we denote the likelihood for n observations $\mathbf{z} = (z_1, \cdots, z_n)$ by $f(\mathbf{z}; \psi)$ where $\psi = (\phi, \theta, \mu, \sigma^2)$ with $\phi = (\phi_1, \cdots, \phi_p)$ and $\theta = (\theta_1, \cdots, \theta_q)$. We introduce as latent variables the unobserved history $\mathbf{y}_0 = (y_0, y_{-1}, \cdots, y_{1-p})$ and $\epsilon_0 = (\epsilon_0, \epsilon_{-1}, \cdots, \epsilon_{1-q})$

resulting in the augmented parameter vector $\psi^* = (y_0, \epsilon_0, \phi, \theta, \mu, \sigma^2)$ of dimension $2(p+q+1)$. The conditional likelihood is then obtained from the factorization

$$f(z|\psi^*) = f(z_1|\psi^*)\, f(z_2|z_1, \psi^*) \cdots f(z_n|z_1, \cdots, z_{n-1}, \psi^*)$$

$$= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\{-\frac{1}{2\sigma^2} \sum_{t=1}^{n} (y_t - \mu_t)^2\} \tag{3}$$

where $$\mu_1 = \sum_{i=1}^{p} \phi_i\, y_{1-i} - \sum_{i=1}^{q} \theta_i\, \epsilon_{1-i},$$

$$\mu_t = \sum_{i=1}^{p} \phi_i\, y_{t-i} - \sum_{i=1}^{t-1} \theta_i (y_{t-i} - \mu_{t-i}) - \sum_{i=t}^{q} \theta_i\, \epsilon_{t-i} \quad \text{for } t = 2, \cdots, q$$

and $$\mu_t = \sum_{i=1}^{p} \phi_i\, y_{t-i} - \sum_{i=1}^{q} \theta_i (y_{t-i} - \mu_{t-i}) \quad \text{for } t = q+1, \cdots, n.$$

2.2 The general AR(p) likelihood

In the special case of an AR(p) process ($\theta(B) \equiv 1$) latent variables are not needed. We have $\psi = (\phi, \mu, \sigma^2)$ and the exact likelihood can be factored, analogous to (3), as

$$f(z|\psi) = (2\pi)^{-\frac{n}{2}} \left[\prod_{t=1}^{n} \sigma_t^2\right]^{-\frac{1}{2}} \exp\left\{-\tfrac{1}{2} \sum_{t=1}^{n} (y_t - \mu_t)^2 / \sigma_t^2\right\}. \tag{4}$$

In (4), $\mu_1 = 0$, $\sigma_1^2 = \sum_{j=1}^{p} \phi_j^2 \gamma_0 + 2 \sum_{1 \le i < j \le p} \phi_i \phi_j \gamma_{j-i} + \sigma^2$, while for $t = 2, \cdots, p$ $\mu_t = \sum_{j=1}^{t-1} \phi_j y_{t-j}$, $\sigma_t^2 = \sum_{j=t}^{p} \phi_j^2 \gamma_0 + 2 \sum_{t \le i < j \le p} \phi_i \phi_j \gamma_{j-i} + \sigma^2$ and for $t = p+1, \cdots, n$, $\mu_t = \sum_{j=1}^{p} \phi_j y_{t-j}$, $\sigma_t^2 = \sigma^2$.

Here γ_k, $k = 0, 1, 2, \cdots$ is the autocovariance of the AR(p) process of lag k. Using $\gamma_0 = \sigma^2/(1 - \phi_1\rho_1 - \cdots - \phi_p\rho_p)$, we can write

$$\sigma_1^2 = \frac{\sigma^2}{1-\phi_1\rho_1-\cdots-\phi_p\rho_p}\left\{\sum_{j=1}^{p}\phi_j^2 + 2\sum_{i<j}\phi_i\phi_j\rho_{j-i}\right\} + \sigma^2$$

and for $t = 2,\cdots,p$

$$\sigma_t^2 = \frac{\sigma^2}{1-\phi_1\rho_1-\cdots-\phi_p\rho_p}\left\{\sum_{j=t}^{p}\phi_j^2 + 2\sum_{i\le t<j}\phi_i\phi_j\rho_{j-i}\right\} + \sigma^2$$

where $\rho_j = \frac{\gamma_j}{\gamma_0}$, so that, for all t, σ_t^2 is of the form $\sigma^2 c_t$. The ρ_j's can be obtained in terms of the ϕ_j's using McLeod's (1975, 1977) algorithm.

3. Complete conditional distributions

Given a prior distribution on ψ^*, $\pi(\psi^*)$ the posterior density for ψ^*

$$\pi(\psi^*|z) \propto f(z|\psi^*)\cdot\pi(\psi^*) \tag{5}$$

Bayesian inference proceeds by obtaining marginal posterior distributions of the components of ψ^* as well as features of these distributions.

The Gibbs sampler introduced by Gelfand and Smith (1990) as a tool for carrying out Bayesian calculations is a Markovian updating scheme which requires sampling from the complete conditional distributions associated with ψ (see e.g., Gelfand and Smith or Gelfand et. al., 1990, for details). A key point is that each complete conditional density is also proportional to the right side of (5). In certain cases, we may recognize this form as that of a standard distribution. In more challenging cases, it emerges only as a non standard, non normalized density. We shall see that, for the Bayesian models in (5), the complete conditional distributions for μ and σ^2 illustrate the former case with ϕ and θ illustrating the latter.

In the general ARMA (p,q) case it is tempting to consider a noninformative prior specification for ψ^*, $\pi(\psi^*) \propto \sigma^{-2}$. Unfortunately such specification yields posterior distributions for y_0 and ϵ_0 which are improper. (This may be readily seen from

straightforward calculations for, say, an ARMA (1,0) model). Rather, since the y_0 and ϵ_0 are, in fact, like y_t and ϵ_t respectively, their prior specification should be a proper distribution given ψ. Hence the prior for ψ^* takes the form

$$\pi(\psi^*) = \pi(y_0, \epsilon_0 \mid \psi) \cdot \pi(\psi)$$

where we can assume a noninformative specification for $\pi(\psi)$, i.e., $\pi(\psi) \propto \sigma^{-2}$.

Indeed, Newbold (1974, p. 424) presents the choice of $\pi(y_0, \epsilon_0 \mid \psi)$ which, upon integration over y_0 and ϵ_0, yields the previously mentioned exact likelihood. It is a multivariate normal with mean 0 and covariance matrix Λ arising from the stationary ARMA process. Working with the $(p+q)\times(p+q)$ matrix Λ presents the same computational problem as working with the exact likelihood. We make a simplification which, on both intuitive and empirical grounds, little affects inference about ψ or forecasts. We replace Λ with $\tilde{\Lambda} = \begin{bmatrix} \gamma_0 I_{p\times p} & 0 \\ 0 & \sigma^2 I_{q\times q} \end{bmatrix}$ where γ_0 is the variance of the stationary ARMA (p,q) process and σ^2 is the assumed error variance. Thus, the joint posterior density for ψ^* is

$$\pi(\psi^* \mid z) \quad \propto \quad (\sigma^2)^{-\frac{(n+2)}{2}} \exp\{-\frac{1}{2\sigma^2} \sum_{t=1}^{n} (y_t - \mu_t)^2\} \, \pi(y_0, \epsilon_0 \mid \psi) \qquad (6)$$

with μ_t defined below (3).

With a little manipulation, we may show that the complete conditional distribution for μ is normal with mean $\frac{1}{n} \sum_{t=1}^{n} (z_t - \mu_t)$ and variance $\frac{\sigma^2}{n}$ and that for σ^2 it is inverse Gamma, i.e., $IG(\frac{n}{2}, \frac{1}{2} \sum_{t=1}^{n} (y_t - \mu_t)^2)$. We note that a constant mean μ for the z's need not be assumed. Rather, for observation z_t we could replace μ with $x^T\beta$ where x is a vector of covariates. No complications to the sampling–based approach result; the normal complete conditional distribution for μ is replaced by a multivariate normal complete conditional distribution for β. The complete conditional densities for the ϕ_j's and θ_j's are proportional to (6) and must be sampled subject to the restriction to $C_p \times C_q$. We develop an efficient

sampling procedure in Section 5. The complete conditional distributions of the latent variables ϵ_{1-j} and y_{1-j} can be shown to all be normal. However calculation of means and variances requires complicated bookkeeping as the expressions involve summations over $t = 1, 2, \cdots, n$ with individual terms having forms which change according to the values of t, j, p and q. To ameliorate the programming burden we have found it easier to again just work with (6).

In the purely autoregressive case, AR(p), the noninformative prior $\pi(\phi, \mu, \sigma^{-2}) \propto \sigma^2$ can be used. From (4) we can see that μ is conditionally normal with mean

$$\left[\sum_{t=1}^{n} \frac{1}{\sigma_t^2}\right]^{-1} \sum_{t=1}^{n} \frac{(z_t - \mu_t)}{\sigma_t^2} \text{ and variance } \left[\sum_{t=1}^{n} \frac{1}{\sigma_t^2}\right]^{-1}$$

and that

$$f(\sigma^2 \mid \phi, \mu, z) \propto (\sigma^2)^{\frac{-(n+2)}{2}} \exp\left\{-\frac{1}{\sigma^2}\left[\frac{1}{2}\sum_{t=1}^{n}\left[\frac{y_t - \mu_t}{c_t}\right]^2\right]\right\}$$

which is an inverse Gamma density, where the c_t are defined below (4). However, as in the general ARMA case, sampling the ϕ_j subject to the stationarity restrictions is not routine.

4. A useful reparametrization

Consider, first, the pure autoregressive model. The constrained parameter region for ϕ, C_p, is analytically intractable for p>4. To complicate matters further, the Gibbs sampler requires cross–sections of C_p, that is, sets for ϕ_j given ϕ_i, $i \neq j$. To circumvent the problem of dealing with these parameter constraints, we consider successive transformations from C_p to a p–dimensional hypercube and then to R^p. In particular, Barndorff–Nielsen and Schou (1973) reparametrize ϕ in terms of the partial autocorrelations r_j of the AR(p) process. This transformation which is one–to–one is defined by

$$\phi_k^{(k)} = r_k$$
$$\phi_i^{(k)} = \phi_i^{(k-1)} - r_k \phi_{k-i}^{(k-1)} \quad i = 1,\cdots,k-1 \tag{7}$$

where $\phi = (\phi_1^{(p)},\cdots,\phi_p^{(p)})$, $\phi_j^{(p)}$ being the j^{th} coefficient from an AR(p) process and $r = (r_1,\cdots,r_p)$. Monahan (1984) reports the inverse transformation in the iterative form

$$\phi_i^{(k-1)} = (\phi_i^{(k)} + \phi_k^{(k)} \phi_{k-i}^{(k)})/(1-(\phi_k^{(k)})^2) \quad i=1,2,\cdots k-1 \tag{8}$$

and the Jacobian

$$J = \prod_{k=1}^{p} (1-r_k^2)^{[(k-1)/2]} \prod_{j=1}^{[p/2]} (1-r_{2j}) \tag{9}$$

The condition that $\phi \in C_p$ now becomes $|r_k|<1$. We, then apply a second transformation from r to $r^* \in \mathbb{R}^p$. Based on the experience of other authors, for example, Smith et. al. (1987), we use the "Fisher–type" transformation discussed in Marriott and Smith (1992) which performs best for densities with "normal shapes" or mixtures of such shapes:

$$r_j^* = \log\left[\frac{1+r_j}{1-r_j}\right] \quad j=1,\cdots,p.$$

We would do all of the random generation in the space of the r_j^*'s inverting the r^*'s back to ϕ's at the end.

The complete conditional density for r_j^* arises by transforming the nonnormalized joint density (5) from (ϕ, μ, σ^2) to (r^*, μ, σ^2) and considering the resulting expression as a function of r_j^* with r_i^*, $i\neq j$, μ and σ^2 fixed. We denote this nonnormalized complete conditional form by $f(r_j^* | r_i^*, i\neq j, \mu, \sigma^2)$.

For the general ARMA model, assuming invertibility and stationarity, we must, in addition, generate θ_i's from their complete conditional distributions, restricting θ to C_q. Monahan (1984) has shown that the above two–stage transformation can also be used for the moving average parameters. We merely replace ϕ with θ and p with q. Applying the transformation to both ϕ and θ results in a transformation from $(\phi,\theta) \in C_p \times C_q$ to say $(r_\phi^*, r_\theta^*) \in R^{p+q}$.

5. Efficient sampling; a modified Gibbs sampler

The standard Gibbs sampler (see Section 3) proceeds by making draws from the complete conditional distributions in some systematic order. When these distributions are familiar forms sampling is routine. When this is not so (and when the complete conditional density does not have special properties such as log concavity) a variety of general techniques have been proposed. These include approximate c.d.f. inversion (as in Tanner, 1991), ratio–of–uniforms (as in Wakefield, Gelfand and Smith, 1991) and adaptive mixtures (as in West, 1992).

An alternative which we employ here, for each such nonroutine draw, runs a scalar valued Markov chain Monte Carlo algorithm whose stationary distribution is the desired complete conditional. Müller (1991) proves that, under mild conditions, use of such univariate trajectories within a trajectory of the Gibbs sampler results in a Markov chain whose stationary distribution is the desired joint posterior distribution.

More precisely, in our present case suppose we use Ansley's (1979) algorithm to obtain the MLE for ϕ and θ say $(\hat{\phi}, \hat{\theta})$ with associated asymptotic covariance matrix $\Sigma_{(\hat{\phi}, \hat{\theta})}$. Using the transformation of Section 4 we convert $(\hat{\phi}, \hat{\theta})$ to, say, $(\hat{r}^*_\phi, \hat{r}^*_\theta)$ and, via the delta method obtain an approximation to the covariance matrix of $(\hat{r}^*_\phi, \hat{r}^*_\phi)$ which we denote by Σ^*. Let $g_{p+q}(r^*_\phi, r^*_\theta)$ be the multivariate normal distribution over R^{p+q} having mean $(\hat{r}^*_\phi, \hat{r}^*_\theta)$ and covariance matrix Σ^*. For any coordinate r^*_j (from either r^*_ϕ or r^*_θ) let $g(r^*_j | r^*_{(-j)})$ denote its univariate conditional normal distribution given all of the other r^*'s derived from g_{p+q}.

We now take $g(r^*_j | r^*_{(-j)})$ to be a Gaussian proposal for a univariate Metropolis algorithm. That is, we create the following Markov chain. If the current value of $r^*_j = u$ and a draw from $g(r^*_j | r^*_{(-j)})$ yields v, we calculate the ratio $\alpha(u,v) = f(v | r^*_i, i \neq j, \mu, \sigma^2)/f(u | r^*_i, i \neq j, \mu, \sigma^2)$. If $\alpha \geq 1$ we move to v; if $\alpha < 1$ we move to v

with probability α. It may be straight forwardly demonstrated that the stationary distribution of this Markov chain is the normalized density associated with $f(r_j^* | r_i^*, i \neq j, \mu, \sigma^2)$ (see, e.g., Hastings, 1970 or Tierney, 1991). An iteration of this modified Gibbs sampler is thus implemented by, starting with r_1^*, running the associated Markov chain for m steps, taking the state of r_1^* at the m^{th} step as the updated value and then proceeding to r_2^*, etc. After all the r_j^* have been updated we draw μ, then σ^2, then y_0 and then ϵ_0 to complete one iteration.

This hybrid algorithm is attractive in our case because the f's tend to be unimodal and roughly normal making the Gaussian proposal we are using efficient. Empirical experience has shown that this algorithm is readily automated and requires far fewer likelihood evaluations than the previously mentioned techniques resulting in substantially shorter run times.

To initiate independent parallel replication of the sampler we perturb the MLE ($\hat{r}_\phi^*$, $\hat{r}_\theta^*$, $\hat{\mu}$, $\hat{\sigma}^2$) obtaining a total of m_1 starting values (we use $m_1 = 40$). After proceeding for k_1 iterations with these m_1 replications, by resampling we increase m_1 to m_2 replications (we use m_2=200). We proceed for k_2 further iterations with the m_2 replications, monitoring the stability of selected quantiles to judge when convergence may be assumed. Finally we increase m_2 to m_3 replications (we use $m_3 = 500$). The sampler is then run for k_3 further iterations before termination. See Gelfand and Smith, (1990) and Gelfand et al. (1990) for further discussion of such sampling schedules.

In concluding this section we mention an alternative approach for sampling the ϕ's and θ's based upon an idea in Jones (1987). Jones observes that, for instance, ϕ may be drawn uniformly over C_p by drawing r according to a product Beta distribution over the hypercube $|r_j| < 1$, $j=1,2,\cdots,p$. Using ideas in Smith and Gelfand (1992), for given μ and σ^2, such ϕ's may then be resampled to essentially have the distribution $f(\phi | \mu, \sigma^2, z)$. However, this resampling would have to be done for each iteration within each replication

of the Gibbs sampler (i.e. for each new μ, σ^2 pair). We believe that this method would be a very inefficient competitor to our sampling scheme.

6. Missing Values and Outliers

In this section, we propose the use of the sampling based approach to handle missing values and to detect outliers in an ARMA(p,q) process. This is done by treating such a value, say z_r, as an unknown and adding it as a parameter to the Gibbs sampler. The complete conditional distribution for z_r, $f(z_r | \mathbf{z}_{-r}, \psi^*)$, is proportional to the right side of (6) and in fact is a normal distribution with mean and variance having forms similar to those for the y_{1-j}. Alternatively, a convenient Gaussian proposal is the conditional distribution $f(z_r | z_1, \cdots, z_{r-1}, \psi^*)$ in (3). A natural candidate for a starting value $z_r^{(0)}$ is the mean of the available z_i or perhaps an interpolated value using adjacent observed z's. We perturb $z_r^{(0)}$ to obtain m_1 starting values for the m_1 initiations of the sampler. An iteration of the Gibbs sampler then proceeds as discussed in Section 5 generating, in addition, a z_r. It is straightforward to extend this procedure to the situation with several missing values or outliers.

7. Prediction

In a Bayesian analysis of ARMA models, prediction proceeds via the predictive density

$$f(\mathbf{z}_F | \mathbf{z}) = \int f(\mathbf{z}_F | \mathbf{z}, \psi^*) \cdot \pi(\psi^* | \mathbf{z}) \, d\psi^* \qquad (10)$$

where $f(\mathbf{z}_F | \psi^*, \mathbf{z})$ is the density of the future data $\mathbf{z}_F$. If $\mathbf{z}_F = (z_{n+1}, \cdots, z_{n+L})$ then $f(\mathbf{z}_F | \mathbf{z}, \psi^*) = f(z_{n+1} | \mathbf{z}, \psi^*) \cdot f(z_{n+2} | z_{n+1}, \mathbf{z}, \psi^*) \cdots f(z_{n+L} | z_{n+1}, \cdots, z_{n+L-1}, \mathbf{z}, \psi^*)$. The use of (10) contrasts sharply with the non–Bayesian practice in time series analysis of basing forecasts on a particular set of estimated parameter values, i.e., on $f(\mathbf{z}_F | \mathbf{z}, \hat{\psi}^*)$. It is well

known (see e.g. Carlin and Gelfand (1990)) that although point estimates arising from $f(z_F | z, \hat{\psi}^*)$ perform well, interval estimates will generally be too short, the variance of this distribution too small. In fact there is recent literature on correcting for this effect by incorporating the variability due to parameter estimation into the prediction mean square error (Stine, 1987).

In studying (10) the sampling based approach again fits nicely. Using the output of the Gibbs Sampler, ψ_j^* , j=1,$\cdots$,m, an approximation to the density itself may be obtained as the Monte Carlo integration

$$\hat{f}(z_F | z) = \frac{1}{m} \sum_{j=1}^{m} f(z_F | z, \psi_j^*). \tag{11}$$

To obtain a sample of predictions from the density (10), for each ψ_j^* we draw $z_{F,j}$ from $f(z_F | z, \psi_j^*)$.

8. Model determination

Model determination involves two aspects – model choice (selection amongst models) and model adequacy (performance of a particular model). Bayesian assessment of model adequacy resides in predictive distributions by which comparison is made between what the model predicts and what was observed. (In fact most any model evaluation scheme relies on such comparison).

For pairwise choice between models the formal Bayes criterion requires calculation of the Bayes factor (ratio of predictive distributions) adjusted by a weight which can be regarded as the prior odds associated with the models. Poskitt and Tremayne (1983) nicely unify such Bayesian model selection for time series models in showing that various established criteria, e.g., AIC, BIC, ϕ_c and S may be viewed as approximate Bayes factors adjusted by weights reflecting sample size and model dimension. Unfortunately, with improper priors on the parameters (as in our case) predictive distributions become improper so that interpretation of the Bayes factor is unclear. Possible remedies involve

utilizing the predictive distribution in a different form. See Gelfand, Dey and Chang (1992) for a recent discussion as well as for ad hoc criteria compatible with the sampling–based approach we have adopted here.

Perhaps more importantly, we would prefer that model selection not be based upon a single number. Rather, we would prefer to compare the predictive performance of competing models at each time point. Hence we propose to look at the predictive distribution for the entire series given what we have observed. More specifically if the observed series is $\mathbf{z}_{obs} = (z_{1,obs}, \cdots, z_{n,obs})$ and we wish to predict a replication of the series $\mathbf{z} = (z_1, \cdots, z_n)$ given $\mathbf{z}_{obs}$ (and a particular ARMA model) then the predictive distribution of $\mathbf{z}$ given $\mathbf{z}_{obs}$ is, analogous to (10),

$$f(\mathbf{z}|\mathbf{z}_{obs}) = \int f(\mathbf{z}|\psi^*)\, \pi(\psi^*|\mathbf{z}_{obs}) \tag{12}$$

To draw samples from $f(\mathbf{z}|\mathbf{z}_{obs})$ we need only draw ψ^* from $\pi(\psi^*|\mathbf{z}_{obs})$ and then, given ψ^*, draw $\mathbf{z}$ from $f(\mathbf{z}|\psi^*)$. The output of our modified Gibbs sampler provides a sample ψ_j^*, $j=1, \cdots, m$ from the posterior. Given ψ_j^*, $\mathbf{z}_j$ can be drawn sequentially using (3).

Comparison of the sampled $\mathbf{z}_j$'s, $j=1, \cdots, m$ with $\mathbf{z}_{obs}$ can be done in many ways. We view such comparison in a diagnostic fashion eschewing formal inference. In particular an appealing graphical display, which is used for the examples of Section 9, may be created as follows. If $\mathbf{z}_j \sim f(\mathbf{z}|\mathbf{z}_{obs})$ then the t^{th} component of $\mathbf{z}_j$, $z_{j\,t} \sim f(z_t|\mathbf{z}_{obs})$ the predictive distribution at the t^{th} time point. Suppose we use the sample $\{z_{j\,t}, j=1, \cdots, m\}$ to obtain $E(z_t|\mathbf{z}_{obs})$ and $var(z_t|\mathbf{z}_{obs})$ and plot $|z_{t,obs} - E(z_t|\mathbf{z}_{obs})|$ vs $\sqrt{var(z_t|\mathbf{z}_{obs})}$.

This display reveals model adequacy via a point cloud close to the origin, i.e., predictive distributions have small dispersion and the observations are consonant with these distributions. Furthermore, all points lying below the line $y=2x$ on this plot are such that the observed y roughly falls within a 95% predictive interval. Outlying observations will lie well above this line. Extending these ideas, the display becomes an informal model

choice plot. In particular, suppose we are comparing nested models. As we run through a portfolio of nested models, with increasing dimensionality, we expect to pass from say a poor model to a satisfactory parsimonious choice to an overfitted one. This will be manifested in the display as follows. The poor model will perform badly on both the x and y scales, the satisfactory parsimonious model will perform well on both scales, the overfitted model will perform well on the y scale but will yield less concentrated predictive distributions than the parsimonious one and hence do worse on the x scale. The same principles should guide choice between two nonnested models, e.g., ARMA (3,5) vs. ARMA (4,2) – the one with the point cloud closer to the origin is preferred. When the two clouds overlap considerably choice between models is unclear. This recognition seems preferable to a decision based upon a single number. Returning to nested models, an alternative informal model selection approach looks at the posterior distribution of the "discrepancy" parameters, i.e., the parameters in the full model which are not in the reduced model. For instance, in comparing an ARMA (3,3) with an ARMA (2,3), we would examine the posterior distribution of ϕ_3 by using the generated ϕ_{3j}, j=1, $\cdots$, m to see where 0 falls. Again, we illustrate the use of these model determination tools in conjunction with the examples of the next section.

9. Illustrative Examples

We present two examples to illustrate our methodology. Example 1 consists of the quarterly seasonally adjusted U.S. unemployment rate between 1948–1972 (Fuller, 1976), a series of n=100 observations. We model this data by autoregressive processes. AR(p) models, $1 \leq p \leq 5$ have been fitted by Shumway (1988), and the AR(2) model was selected as a best parsimonious choice since the maximum likelihood estimates of ϕ_3, ϕ_4 and ϕ_5 are not significantly different from 0. Using the likelihood in (4) we present here the results from an exact Bayesian analysis. In Table 1 we fit the AR(1) model based on the first 96 observations, the last 4 observations being held out for forecast evaluation. The maximum

likelihood estimates for the parameters, from which the Gibbs sampler is started, are obtained by the Ansley algorithm. Table 1 presents the MLE's together with their standard errors. Also given are the mean and standard derivation as well as selected percentiles of the marginal posterior distributions for the parameters obtained from the Gibbs sampler. We note that the values at the 50^{th} percentile for the ϕ_1, μ and σ^2 correspond closely to the MLE's. The extreme percentiles enable equal tail Bayesian interval estimates.

Similarly, Tables 2–5 present the results of fitting the AR(2), AR(3), AR(4) and AR(5) models respectively. Interval estimates for the parameters ϕ_3, ϕ_4 and ϕ_5 comfortably contain 0, enabling selection of the AR(2) model. For this model we also present in Table 2, results from assuming 0, 1, 5 and 20 percent of the n=96 observations as missing. As in Section 6, we treat the missing observations as parameters along with ϕ_1, ϕ_2, μ, σ^2. Inference seems to be little affected even with as much as 20% missing data. Table 6 presents the forecasts for the data modeled by AR(2) model, as described in Section 7, for t=97, 98, $\cdots$, 100. Figure 1 presents a display of the type described in Section 8 showing the AR(1), AR(2) and AR(5) models. Clearly the AR(1) model is poor with the AR(2) and AR(5) models quite similar.

The data for the second example are the logarithms (base 10) of the Canadian lynx trap counts over a 114 year period (1821–1934). This series has been modeled in the literature by, e.g., Priestley (1981) and Tong (1977). In the class of ARMA models, the AR(2), AR(11) and ARMA (3,3) models have been discussed. Note that there is no nesting between the latter two. The fits for the AR(2) and AR(11) were obtained analogously to those in the first example and are presented in Tables 7 and 8 respectively. The ARMA (3,3) model was fit using the likelihood in (6) with results given in Table 9. Finally Figure 2 presents a display of the type discussed in Section 8. The AR(11) appears preferable to the AR(2) but the AR(3,3) seems the best of the three.

References

Ansley, C.F. (1979). An algorithm for the exact likelihood of a mixed autoregressive moving average process. *Biometrika, 66*, 59–65.

Barndorff–Neilsen, O.E., and Schou, G. (1973). On the reparametrization of autoregressive models by partial autocorrelations. *Journal Multivar. Analysis, 3*, 408–19.

Box, G.E.P. and Jenkins, G.M. (1976). Time Series: Analysis, forecasting and control. Holden Day: San Fransisco.

Broemeling, L.D. (1985). Bayesian analysis of linear models. Marcel Dekker: New York.

Broemeling, L.D., and Shaarawy, S. (1988). Time Series: *a* Bayesian analysis in the time domain, in Bayesian analysis of time series and dynamic models, Ed. J.C. Spall, Marcel Dekker.

Carlin, B. and Gelfand, A. (1990). Approaches for Empirical Bayes Confidence Intervals. *Journal American Statistical Association 85*, p. 105–114.

Carlin, B.P, Polson, N.G., and Stoffer, D.S. (1992). A Monte Carlo approach to nonnormal and nonlinear state space modeling. *Journal American Statistical Association* (to appear).

Chib, S. (1991). Bayes regression with autoregressive errors: a Gibbs sampling approach. *Journal of Econometrics* (to appear).

Chib, S. and Greenberg, E. (1992). Bayes inference via Gibbs sampling in regression models with AR(p) and MA(q) errors, Technical Report, Washington University, St. Louis, Mo.

Fuller, W.A. (1976). Introduction to Statistical Time Series. Wiley, New York.

Gelfand, A.E., Dey, D.K. and Chang H. (1992). Model determination using predictive distributions with implementation via sampling–based methods. *Bayesian Statistics 4*, Eds. J. Bernardo, Oxford Univ. Press. (To appear).

Gelfand, A.E. and Smith, A.F.M. (1990), Sampling based approaches to calculating marginal densities, *Journal American Statistical Association, 85*, 398–409.

Gelfand, A.E. Hills, S.E., Racine–Poon, A. and Smith, A.F.M. (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling, *Journal American Statistical Association, 85*, 972–985.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence, 6*, 721–741.

Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications, *Biometrika 57*, 97–19.

Jones, M.C. (1987). Randomly choosing parameters from the stationarity and invertibility region of autoregressive moving average models. *Applied Statistics, 36*, 134–138.

Marriott, J.M. and Smith, A.F.M. (1992). Reparametrization aspects of numerical Bayesian methods for autoregressive moving average models. Journal of Time Series Analysis (to appear).

McCulloch, R.E. and Tsay, R.S. (1991). Bayesian analysis of autoregressive time series via the Gibbs sampler. Univ. of Chicago, School of Business, Tech. Rpt.

McLeod (1975 and 1977). The derivation of the theoretic autocovariance function of autoregressive moving average time. *Applied Statistics, 24*, 255–256 and correction is 26, pg. 194.

Monahan, J.F. (1983). Fully Bayesian analysis of ARMA time series models. *Journal of Econometrics, 21*, 307–331.

Monahan, J.F. (1984). A note on enforcing stationarity in autoregressive moving average models. *Biometrika, 71*, 403–404.

Muller, P. (1991). A generic approach to posterior integration and Gibbs sampling Purdue University, Department of Statistics, Tech. Rpt. 91–09.

Newbold, P. (1974). The exact likelihood for a mixed autoregressive moving average process. *Biometrika, 61*, 423–426.

Poskitt, D.S. and Tremayne, A.R. (1983). On the posterior odds of time series models, *Biometrika 70*, 157–162.

Priestley, M.B. (1981). Spectral analysis and time series. Academic Press, New York.

Shumway, R.H. (1988). Applied Statistical Time Series Analysis. Prentice Hall, New Jersey.

Smith, A.F.M. and Gelfand, A.E. (1992). Bayesian statistics without tears: a sampling–resampling perspective. *American Statistician.* (To appear).

Smith A.F.M., Skene, A.M., Shaw, J.E.H. and Naylor, J.C. (1987). Progress with numerical and graphical methods for practical Bayesian statistics. *The Statistician, 96*, 75–82.

Stine, R.A. (1987). Estimating properties of autoregressive forecasts, *Journal American Statistical Association 82*, 1072–1078.

Tanner, M. and Wong, W. (1987). The calculation of posterior distributions by Data Augmentation, (with discussion) *Journal American Statistical Association, 82*, 528–550.

Tierney, L. (1991). Markov Chains for exploring Posterior Distributions, University of Minnesota, School of Statistics, Tech. Rpt. #560.

Tong, H. (1977). Some comments on the Canadian lynx data. *J. Royal Statist. Soc. A*, 140, 432–36.

Wei, W.W.S. (1990). Time Series Analysis: Univariate and multivariate methods. Addison–Wesley: California.

West, M. (1992). Bayesian computations: Monte–Carlo density estimation, *J. Royal Statist. Soc. B.*, (to appear).

Zellner, A. (1971). An introduction to Bayesian Inference in Econometrics. John Wiley: New York.

Table 1: Parameter Estimates for U.S. Unemployment Data [AR(1) Model]

Parameter	MLE (s.e)	Posterior Features from Gibbs Sampler							
		Mean (s.d)	Percentiles						
			2.5%	5%	25%	50%	75%	95%	97.5%
ϕ_1	.92 (.04)	.95 (.03)	.88	.89	.93	.95	.97	1.00	1.00
μ	4.76 (.53)	5.14 (.17)	4.86	4.89	5.03	5.14	5.24	5.42	5.52
σ	.45	.45 (.03)	.39	.40	.43	.45	.47	.51	.52

Table 2: Parameter Estimates for U.S. Unemployment Data [AR(2) Model]

Parameter	Percentage of Missing Observations	MLE (s.e)	Posterior Features from Gibbs Sampler							
			Mean (s.d.)	Percentiles						
				2.5%	5%	25%	50%	75%	95%	97.5%
ϕ_1	0	1.56 (.07)	1.58 (.07)	1.44	1.46	1.53	1.58	1.62	1.68	1.69
	1		1.57 (.07)	1.43	1.45	1.52	1.57	1.62	1.67	1.69
	5		1.57 (.08)	1.42	1.45	1.52	1.57	1.62	1.70	1.72
	20		1.55 (.08)	1.40	1.43	1.51	1.56	1.61	1.68	1.70
ϕ_2	0	−.69 (.07)	−.70 (.07)	−.83	−.80	−.75	−.71	−.66	−.59	−.57
	1		−.70 (.07)	−.82	−.80	−.75	−.70	−.65	−.59	−.57
	5		−.70 (.08)	−.85	−.83	−.75	−.70	−.65	−.59	−.56
	20		−.69 (.08)	−.86	−.82	−.74	−.70	−.64	−.56	−.54
μ	0	4.74 (.25)	4.79 (.07)	4.65	4.67	4.75	4.80	4.84	4.90	4.92
	1		4.78 (.07)	4.65	4.66	4.73	4.77	4.82	4.88	4.91
	5		4.80 (.07)	4.65	4.67	4.75	4.79	4.84	4.90	4.92

Table 2 – continued

Parameter	Percentage of Missing Observations	MLE (s.e)	Posterior Features from Gibbs Sampler							
			Mean (s.d.)	Percentiles						
				2.5%	5%	25%	50%	75%	95%	97.5%
μ	20		4.79 (.07)	4.65	4.67	4.74	4.79	4.83	4.89	4.91
σ	0	.32	.32 (.03)	.28	.28	.31	.32	.34	.37	.37
	1		.32 (.02)	.28	.29	.31	.32	.34	.36	.37
	5		.32 (.02)	.28	.29	.30	.32	.34	.36	.37
	20		.33 (.03)	.29	.29	.31	.33	.34	.38	.39

Table 3: Parameter Estimates for U.S. Unemployment Data [AR(3) Model]

Parameter	MLE (s.e)	Posterior Features from Gibbs Sampler							
		Mean (s.d)	Percentiles						
			2.5%	5%	25%	50%	75%	95%	97.5%
ϕ_1	1.68 (.10)	1.67 (.10)	1.48	1.51	1.61	1.67	1.73	1.82	1.85
ϕ_2	−.95 (.17)	−.92 (.17)	−1.25	−1.20	−1.02	−.92	−.81	−0.66	−.59
ϕ_3	.16 (.10)	.14 (.10)	−.05	−.02	.07	.13	.21	.30	.33
μ	4.75 (.29)	4.81 (.08)	4.67	4.69	4.75	4.81	4.86	4.93	4.95
σ	.32	.32 (.02)	.28	.29	.31	.32	.33	.36	.37

Table 4: Parameter Estimates for U.S. Unemployment Data [AR(4) Model]

Parameter	MLE (s.e)	Posterior Features from Gibbs Sampler							
		Mean (s.d)	Percentiles						
			2.5%	5%	25%	50%	75%	95%	97.5%
ϕ_1	1.65 (.10)	1.65 (.09)	1.45	1.49	1.58	1.64	1.71	1.80	1.84
ϕ_2	−.83 (.20)	−.80 (.18)	−1.17	−1.12	−.92	−.80	−.70	−.51	−.47
ϕ_3	−.05 (.20)	−.05 (.17)	−.38	−.33	−.16	−.06	−.05	.23	.30
ϕ_4	.12 (.10)	.11 (.09)	−.08	−.04	.05	.11	.18	.25	.28
μ	4.75 (.32)	4.83 (.08)	4.68	4.71	4.78	4.83	4.88	4.96	4.98
σ	.32	.32 (.02)	.28	.28	.30	.32	.34	.36	.37

Table 5: Parameter Estimates for U.S. Unemployment Data [AR(5) Model]

Parameter	MLE (s.e)	Posterior Features from Gibbs Sampler							
		Mean (s.d)	Percentiles						
			2.5%	5%	25%	50%	75%	95%	97.5%
ϕ_1	1.65 (.10)	1.64 (.10)	1.43	1.47	1.58	1.64	1.71	1.80	1.83
ϕ_2	−.82 (.20)	−.81 (.18)	−1.16	−1.12	−.94	−.80	−.70	−.53	−.47
ϕ_3	−.02 (.22)	−.05 (.18)	−.28	−.25	−.07	.04	.16	.35	.40
ϕ_4	.06 (.20)	−.08 (.18)	−.40	−.36	−.19	−.09	.03	.23	.28
ϕ_5	.04 (.10)	.12 (.10)	−.11	−.07	.06	.12	.18	.27	.28
μ	4.75 (.34)	4.85 (.09)	4.67	4.70	4.79	4.85	4.91	5.01	5.03
σ	.32	.32 (.02)	.27	.28	.30	.32	.33	.36	.37

Table 6: Forecasts for US Unemployment Data [AR(2) Model]

Lead	Percentage of Missing Observations	Actual	Based on MLE (s.e)	Percentiles from Gibbs Sampler						
				2.5%	5%	25%	50%	75%	95%	97.5%
1	0	5.83	5.81 (0.35)	5.13	5.31	5.61	5.83	6.05	6.34	6.44
	1			5.15	5.27	5.59	5.81	6.02	6.35	6.44
	5			5.14	5.25	5.60	5.82	6.05	6.39	6.49
	20			5.04	5.11	5.54	5.86	6.11	6.49	6.60
2	0	5.77	5.56 (0.64)	4.37	4.57	5.21	5.61	6.06	6.55	6.74
	1			4.36	4.63	5.13	5.56	5.95	6.55	6.70
	5			4.34	4.54	5.16	5.57	6.02	6.64	6.76
	20			4.03	4.37	5.15	5.63	6.08	6.74	6.89
3	0	5.53	5.28 (0.88)	3.73	3.95	4.81	5.41	5.96	6.72	6.93
	1			3.66	3.93	4.73	5.25	5.88	6.66	6.90
	5			3.64	3.87	4.73	5.28	5.91	6.74	7.00
	20			3.54	3.78	4.79	5.39	5.97	6.95	7.16
4	0	5.30	5.02 (1.05)	3.09	3.42	4.43	5.09	5.80	6.75	7.14
	1			3.16	3.35	4.26	4.99	5.73	6.68	7.20
	5			2.99	3.42	4.39	5.02	5.82	6.89	7.14
	20			3.02	3.38	4.47	5.18	5.85	6.91	7.27

Table 7: Parameter Estimates for Logarithms of Canadian Lynx Data [AR(2) Model]

Parameter	MLE (s.e)	Posterior Features from Gibbs Sampler							
		Mean (s.d)	Percentiles						
			2.5%	5%	25%	50%	75%	95%	97.5%
ϕ_1	1.38 (.06)	1.39 (.09)	1.19	1.23	1.32	1.39	1.45	1.52	1.53
ϕ_2	−.74 (.06)	−.75 (.09)	−.89	−.87	−.81	−.76	−.69	−.59	−.55
μ	2.90 (.06)	2.91 (.03)	2.85	2.86	2.89	2.91	2.93	2.95	2.97
σ	.23	.23 (.02)	.20	.20	.22	.23	.24	.25	.26

Table 8: Parameter Estimates for Logarithms of Canadian Lynx Data [AR(11) Model]

Parameter	MLE (s.e)	Posterior Features from Gibbs Sampler							
		Mean (s.d)	Percentiles						
			2.5%	5%	25%	50%	75%	95%	97.5%
ϕ_1	1.17 (.09)	1.15 (.09)	.97	1.00	1.08	1.14	1.21	1.28	1.33
ϕ_2	−.55 (.14)	−.51 (.14)	−.80	−.75	−.61	−.51	−.42	−.29	−.26
ϕ_3	.27 (.15)	.24 (.15)	−.04	0.00	.14	.23	.33	.49	.54
ϕ_4	−.31 (.15)	−.29 (.14)	−.58	−.53	−.38	−.28	−.19	−.07	−.02
ϕ_5	.15 (.15)	.13 (.14)	−.14	−.10	.04	.14	.22	.36	.39
ϕ_6	−.15 (.15)	−.12 (.14)	−.42	−.36	−.22	−.12	−.03	−.10	.13
ϕ_7	.06 (.15)	.03 (.14)	−.23	−.19	−.06	.03	.13	.26	.31
ϕ_8	−.03 (.15)	−.01 (.15)	−.30	−.25	−.10	.00	.09	.23	.26
ϕ_9	.14 (.14)	.11 (.14)	−.18	−.13	.02	.11	.21	.35	.39
ϕ_{10}	.20 (.14)	.21 (.13)	−.04	0.00	.12	.21	.29	.43	.47
ϕ_{11}	−.34 (.09)	−.34 (.08)	−.50	−.48	−.40	−.34	−.28	−.21	−.19

Table 8 – continued

Parameter	MLE (s.e)	Posterior Features from Gibbs Sampler							
		Mean (s.d)	Percentiles						
			2.5%	5%	25%	50%	75%	95%	97.5%
μ	2.91 (.05)	2.90 (.02)	2.85	2.86	2.88	2.90	2.92	2.94	2.95
σ	.19	.19 (.01)	.17	.17	.19	.19	.20	.21	.22

Table 9: Parameter Estimates for Logarithms of Canadian Lynx Data [AR(3,3) Model]

Parameter	MLE (s.e)	Posterior Features from Gibbs Sampler							
		Mean (s.d)	Percentiles						
			2.5%	5%	25%	50%	75%	95%	97.5%
ϕ_1	1.99 (.13)	1.86 (.21)	1.36	1.47	1.72	1.88	2.01	2.15	2.20
ϕ_2	−1.15 (.21)	−1.43 (.33)	−1.98	−1.89	−1.68	−1.47	−1.21	−.87	−.68
ϕ_3	.42 (.13)	.29 (.20)	−.16	−.05	.16	.31	.44	.56	.61
θ_1	.83 (.12)	.66 (.22)	.18	.29	.53	.68	.80	.99	1.03
θ_2	−.05 (.13)	.01 (.14)	−.31	−.25	−.07	.03	.11	.22	.25
θ_3	−.53 (.09)	−.49 (.11)	−.69	−.65	−.56	−.49	−.42	−.32	−.27
μ	2.91 (.06)	2.91 (.03)	2.86	2.87	2.90	2.91	2.93	2.96	2.96
σ	.20	.20 (.01)	.18	.18	.20	.20	.21	.23	.23

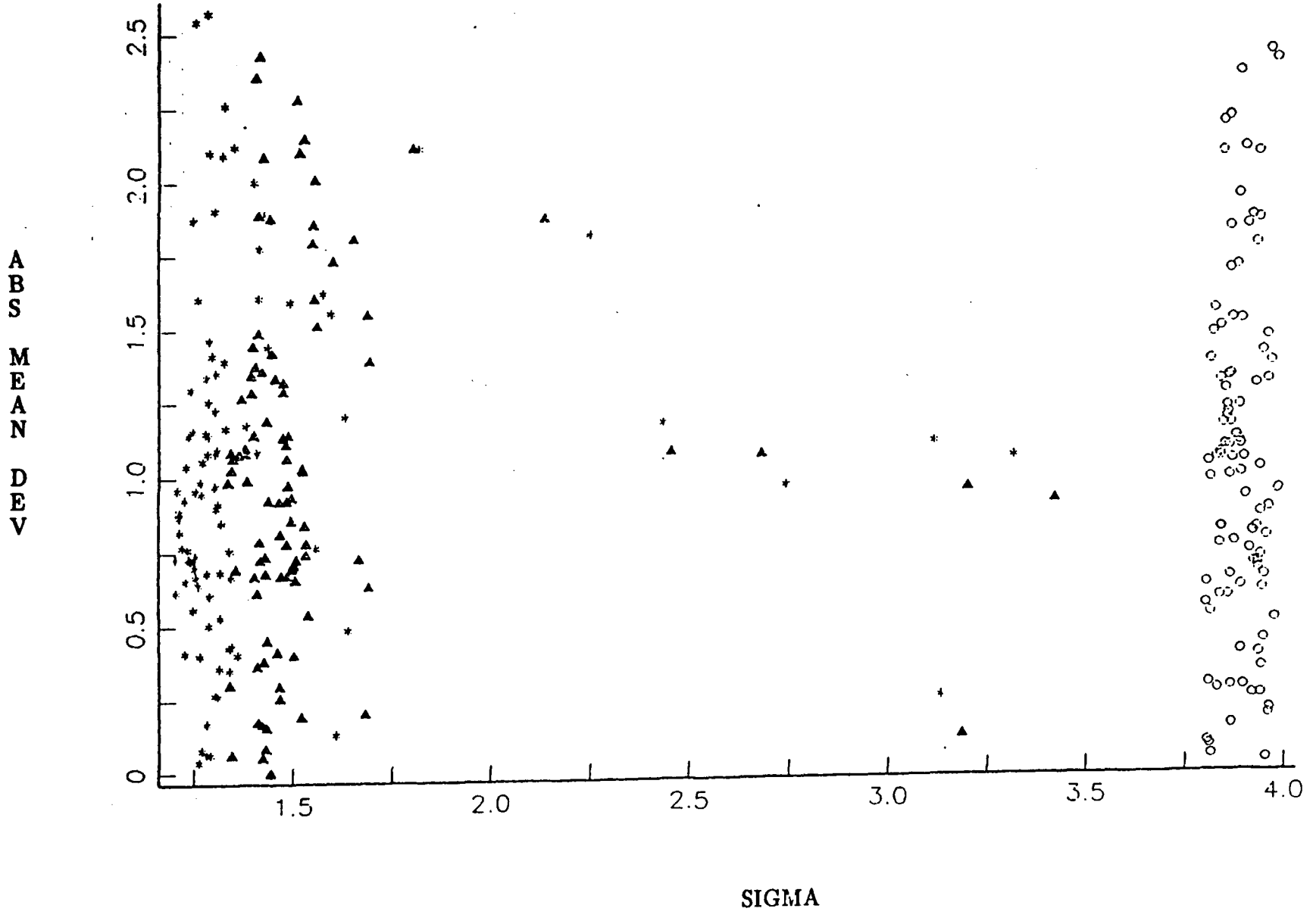

Figure 1: Model choice plot for the U.S. Unemployment Data
0=AR(1), Δ=AR(2), *=AR(5)

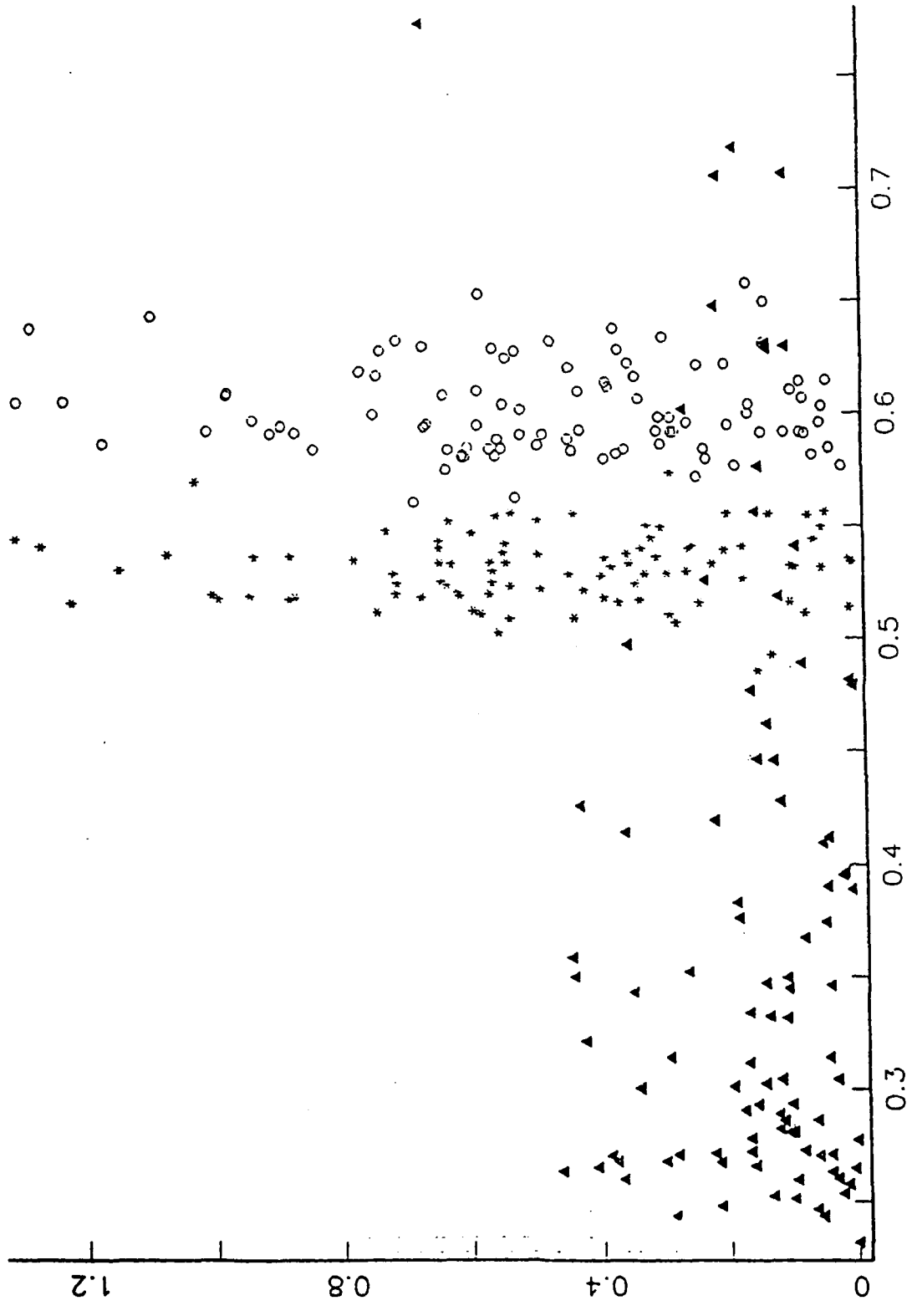

Figure 2: Model choice plot for the Canadian Lynx Data
0=AR(2), Δ=AR(11), *=AR(3,3)

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Bayesian Analysis of ARMA Processes: Complete Sampling Based Inference Under Full Likelihoods		5. TYPE OF REPORT & PERIOD COVERED Technical 6. PERFORMING ORG. REPORT NUMBER No. 471
7. AUTHOR(s) John Marriott, N. Ravishanker, & A.E. Gelfand		8. CONTRACT OR GRANT NUMBER(s) N0025-92-J-1264
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, CA 94305-4065		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-042-267
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics & Probability Program Code 111		12. REPORT DATE June 24, 1993 13. NUMBER OF PAGES 22
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

THE VIEW, OPINIONS, AND/OR FINDINGS CONTAINED IN THIS REPORT ARE THOSE OF THE AUTHOR(S) AND SHOULD NOT BE CONSTRUED AS AN OFFICIAL DEPARTMENT OF THE ARMY POSITION, POLICY, OR DECISION, UNLESS SO DESIGNATED BY OTHER DOCUMENTATION.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Keywords: *Gibbs sampler, invertibility, latent variables, Metropolis algorithm, missing values, outliers, prediction, stationarity.*

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

See Reverse Side

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

BAYESIAN ANALYSIS OF ARMA PROCESSES: COMPLETE SAMPLING BASED INFERENCE UNDER FULL LIKELIHOODS

John Marriott, Nalini Ravishanker, Alan Gelfand

and

Jeffrey Pai*

Abstract

For a general stationary and invertible ARMA (p,q) process, we show how to carry out a fully Bayesian analysis. Our approach is through the use of sampling based methods involving three novel aspects. First the constraints on the parameter space arising from the stationarity and invertibility conditions are handled by a convenient reparametrization to all of Euclidean (p+q)-space. Second, required sampling is facilitated by the introduction of latent variables which, though increasing the dimensionality of the problem, greatly simplifies the evaluation of the likelihood. Third, the particular sampling based approach used is a Markov chain Monte Carlo method which is a hybrid of the Gibbs sampler and the Metropolis algorithm. We also briefly show how straightforwardly the sampling based approach accommodates missing observations, outlier detection, prediction and model determination. Finally we illustrate the approach with two examples.